mBio

ABSTRACT

Authors: Lynda Rocheleau, Geneviève Laroche, Kathy Fu, Corina M. Stewart, Abdulhamid O. Mohamud, Marceline Côté, Patrick M. Giguère, Marc-André Langlois, and Martin Pelchat AUTHORS INFO & AFFILIATIONS

DOI: https://doi.org/10.1128/mBio.00788-21 • Check for updates

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a virus that is continuously evolving. Although its RNA-dependent RNA polymerase exhibits some exonuclease proofreading activity, viral sequence diversity can be produced by replication errors and host factors. A diversity of genetic variants can be observed in the intrahost viral population structure of infected individuals. Most mutations will follow a neutral molecular evolution and will not make significant contributions to variations within and between infected hosts. Herein, we profiled the intrasample genetic diversity of SARS-CoV-2 variants, also known as quasispecies, using high-throughput sequencing data sets from 15,289 infected individuals and infected cell lines. Despite high mutational background, we identified recurrent intragenetic variable positions in the samples analyzed, including several positions at the end of the gene encoding the viral spike (S) protein. Strikingly, we observed a high frequency of C→A missense mutations resulting in the S protein lacking the last 20 amino acids (SΔ20). We found that this truncated S protein undergoes increased processing and increased syncytium formation, presumably due to escaping M protein retention in intracellular compartments. Our findings suggest the emergence of a high-frequency viral sublineage that is not horizontally transmitted but potentially involved in intrahost disease cytopathic effects.

IMPORTANCE The mutation rate and evolution of RNA viruses correlate with viral adaptation. While most mutations do not make significant contributions to viral molecular evolution, some are naturally selected and produce variants through positive selection. Many SARS-CoV-2 variants have been recently described and show phenotypic selection toward more infectious viruses. Our study describes another type

economic burdens. SARS-CoV-2 is an enveloped virus with a nonsegmented, positive-sense, single-stranded viral RNA (vRNA) genome comprised of ~30,000 nucleotides (1, 2). The virus is composed of four main structural proteins, encoded in the last 3′-terminal third of the viral genome: the spike glycoprotein (S), membrane (M), envelope (E), and nucleocapsid (N) (3–5). Attachment to the host receptor angiotensin-converting enzyme 2 (ACE2) is mediated by the S protein expressed on the surface of the virion (6). Following its association, the S protein is cleaved into two separate polypeptides (S1 and S2), which triggers the fusion of the viral particle with the cellular membrane (6, 7). Once inside a cell, its RNA-dependent RNA polymerase (RdRp), which is encoded in the first open reading frame of the viral genome (8), carries out transcription and replication of the vRNA genome. In addition, mRNAs coding for the structural proteins (e.g., S, M, E, and N) are expressed by subgenomic RNAs (8). Once translated, the S, M, and E proteins localize and accumulate at the CoV budding site in the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) (9). One aspect of CoV biology is that CoV virions bud into the lumen of the secretory pathway at the ERGIC and must then traffic through the Golgi complex and anterograde system to be efficiently released from host cells (10). The S protein possesses an endoplasmic reticulum retrieval signal (ERRS) at its carboxy terminus, which is required for trafficking through the ERGIC (11). At this location, the spike protein interacts with the M protein, which has been shown to be essential for accumulation at the ERGIC. The N protein then associates with the viral genome and assembles into virions, which are transported along the endosomal network and released by exocytosis (8). If not retained at ERGIC, the S protein traffics through the Golgi complex and is preactivated by resident proteases prior to reaching the plasma membrane. Here, it can mediate cell fusion between adjacent cells, resulting in the production of multinucleated cells, or syncytia (7, 12, 13).

Transmitted genomes and consensus sequences are only part of the genetic landscape with regard to RNA viruses. Replication of RNA viruses typically produces quasispecies in which the transmitted viral RNA genomes do not exist as a single sequence entity but rather as a population of genetic variants (25). These mutations are most frequently caused by both the error-prone nature of each of their respective viral RdRps and the host RNA editing enzymes, such as APOBECs and ADARs (26). However, the RdRp complex of large RNA viruses, such as coronaviruses, sometimes possesses exonuclease proofreading activity, and consequently, they have lower error rates (25, 27). Quasispecies may sometimes exhibit diminished replicative fitness or deleterious mutations and exert different roles that are not directly linked to viral genomic propagation (28). Mutations that form the intrahost genetic spectrum have been shown to help viruses evade cytotoxic T cell recognition and neutralizing antibodies, rendering these viruses more resistant to antiviral drugs (28). Additionally, these mutations can also be involved in modulating the virulence and transmissibility of the quasispecies (28).

In this study, we focused on assessing intrahost genetic variations of SARS-CoV-2. We analyzed high-throughput sequencing data sets to profile the sequence diversity of SARS-CoV-2 variants within distinct sample populations. We observed high intrahost genetic variability of the viral genome. By comparing variation profiles between samples from different donors and cell lines, we identified highly conserved subspecies that independently and recurrently arose in different data sets and, therefore, in different individuals. We further analyzed the dominant variant SΔ20 in a functional assay and demonstrate that this truncated S protein avoids inhibition caused by M protein and enhances syncytium formation. We provide evidence for the existence of a consistently emerging variant identified across geographical regions that

for each data set, and the nucleotide compositions for each position were compared to the respective consensuses. To reduce the number of variations due to amplification bias and sequencing errors, duplicated reads were combined, and only positions mapped with a sequencing depth of 50 reads and having at least 5 reads with variations compared to the sample consensus were considered. Overall, we identified 301,742 variations from 11,362 samples located on 26,113 positions of the 29,903-nucleotide (nt) SARS-CoV-2 genome. We observed an average of 26.6 ± 132.0 variable nucleotides per sample (ranging from 1 to 5,295 variations/sample) (Fig. 1A).

FIG 1

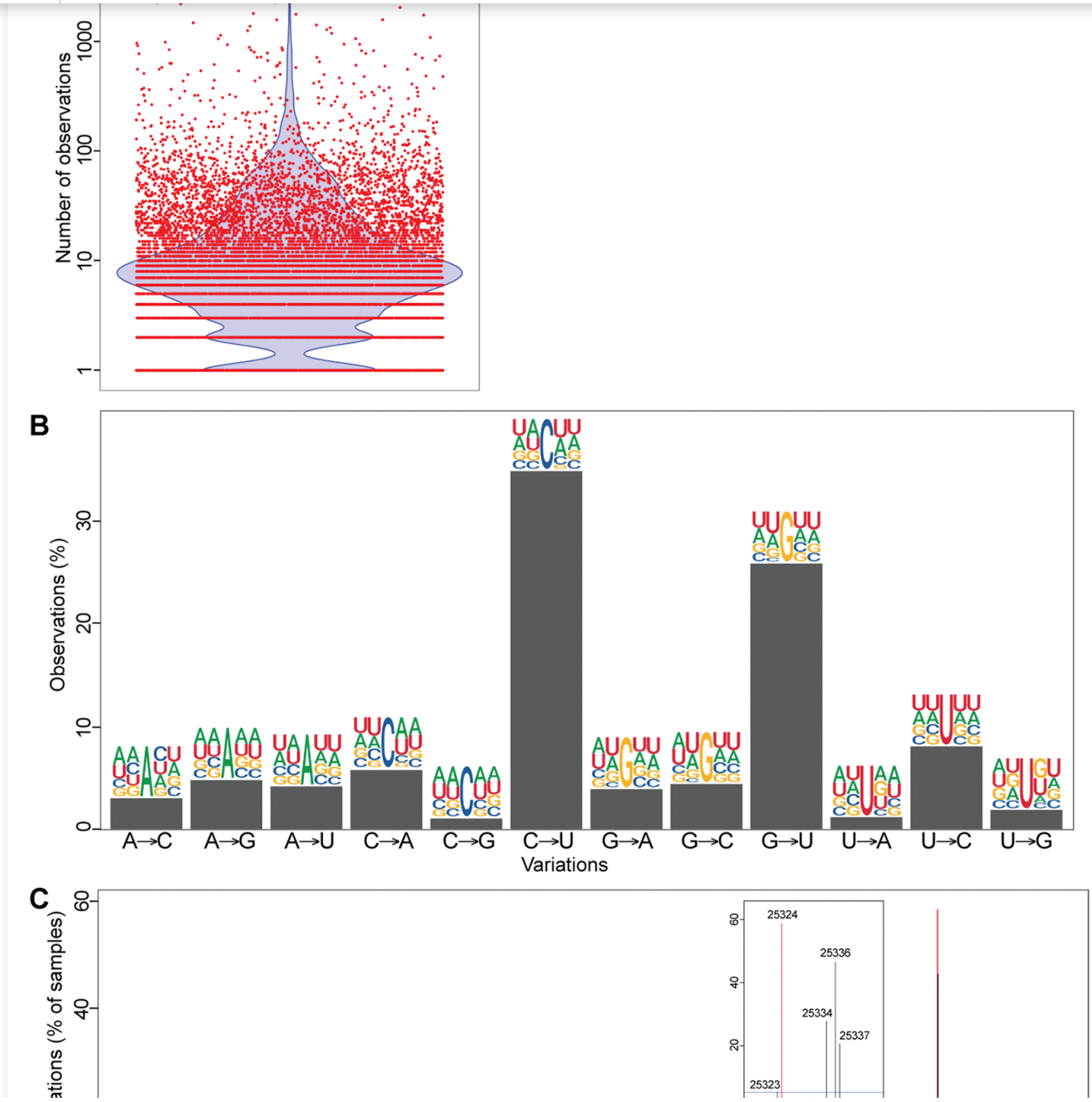


FIG 1 Intrasample variability of the SARS-CoV-2 genome in infected individuals. (A) Number of intragenetic variations observed for each sample analyzed. The red dots represent the 11,362 samples analyzed, and the blue violon shows the distribution of the data. (B) Type of variation and sequence context for each intrasample variable position. Bars represent the percentage of each type. Sequence context is represented by logos comprised of the consensus nucleotides (center) with 2 nt upstream and 2 downstream from each intrasample variable position. (C) Recurrent intragenetic variations are represented as percentages of samples containing variations at each position. The SARS-CoV-2 genome and its genes are represented by yellow boxes below the graph. The blue line indicates 5% shared variations and was used to extract the recurrent intrasample variations listed in Table 1. The inset represents a magnification of the cluster identified at the end of the S gene. (D) One-dimensional representation of the data shown in panel C for each type of variation individually. The location of the C→A variation at position 25,324 is indicated by a red line in panels C and D.

Analysis of the type of intragenetic variations present in SARS-CoV-2 samples from infected individuals.

The analysis of the type of nucleotide changes within samples revealed that 52.2% were transitions (either purine→purine or pyrimidine→pyrimidine) and 47.8% were transversions (purine→pyrimidine or pyrimidine→purine). Notably, the highest nucleotide variations corresponded to C→U transitions (43.5%), followed by G→U transversion (28.1%) (Fig. 1B), both types encompassing 71.6% of all variations. Since editing by host enzymes depends on the sequence context, we extracted 2 nt upstream and downstream from each genomic position corresponding to variations and generated sequence logos. Our results indicated a high number of A's and U's around all variation types and sites (62.1% ± 3.4%) (Fig. 1B).

samples from infected individuals.

To identify biologically relevant intragenetic variations, we examined the variable positions that are recurrent in the samples analyzed. The variable positions were tabulated for each sample, and then recurrent intragenetic variations were calculated as percentages of samples containing a variation at each position. Most variations are distributed homogeneously on the viral genome. The number of variations strongly correlates with the length of each gene (Pearson correlation coefficient of 0.972), and most are poorly shared among samples (Fig. 1C and D). However, our analysis reveals 15 recurrent intragenetic variations shared by at least 5% of the samples analyzed (Fig. 1C, above the blue line; Table 1). Among these, four transversions (at nt 25324, 25334, 25336, and 25337) located at the 3′ end of the S gene are the most recurrent variations (Fig. 1C, inset; Table 1). Three of these transversions (at nt 25334, 25336, and 25337) correspond to missense mutations: E1258D (46.4% of the samples), E1258Q (27.6% of the samples), and D1259H (20.1% of the samples). Interestingly, the most observed variation (at nt 25324) is shared by 58.7% of the samples (6,668 of the 11,362 samples) and corresponds to a C→A transversion producing a nonsense mutation at amino acid 1254 of the S protein (Fig. 1C and D, red lines; Fig. 2A, red rectangle). The resulting S protein lacks the last 20 amino acids (SΔ20), which includes the ERRS motif at its carboxy terminus (Fig. 2A, white letters on a black background). Among the samples with this intragenetic variation, this C→A transversion represents from 2.9 to 42.4% of the subspecies identified (mean of 8.2% ± 2.9%) (Fig. 2B; Table 1).

FIG 2

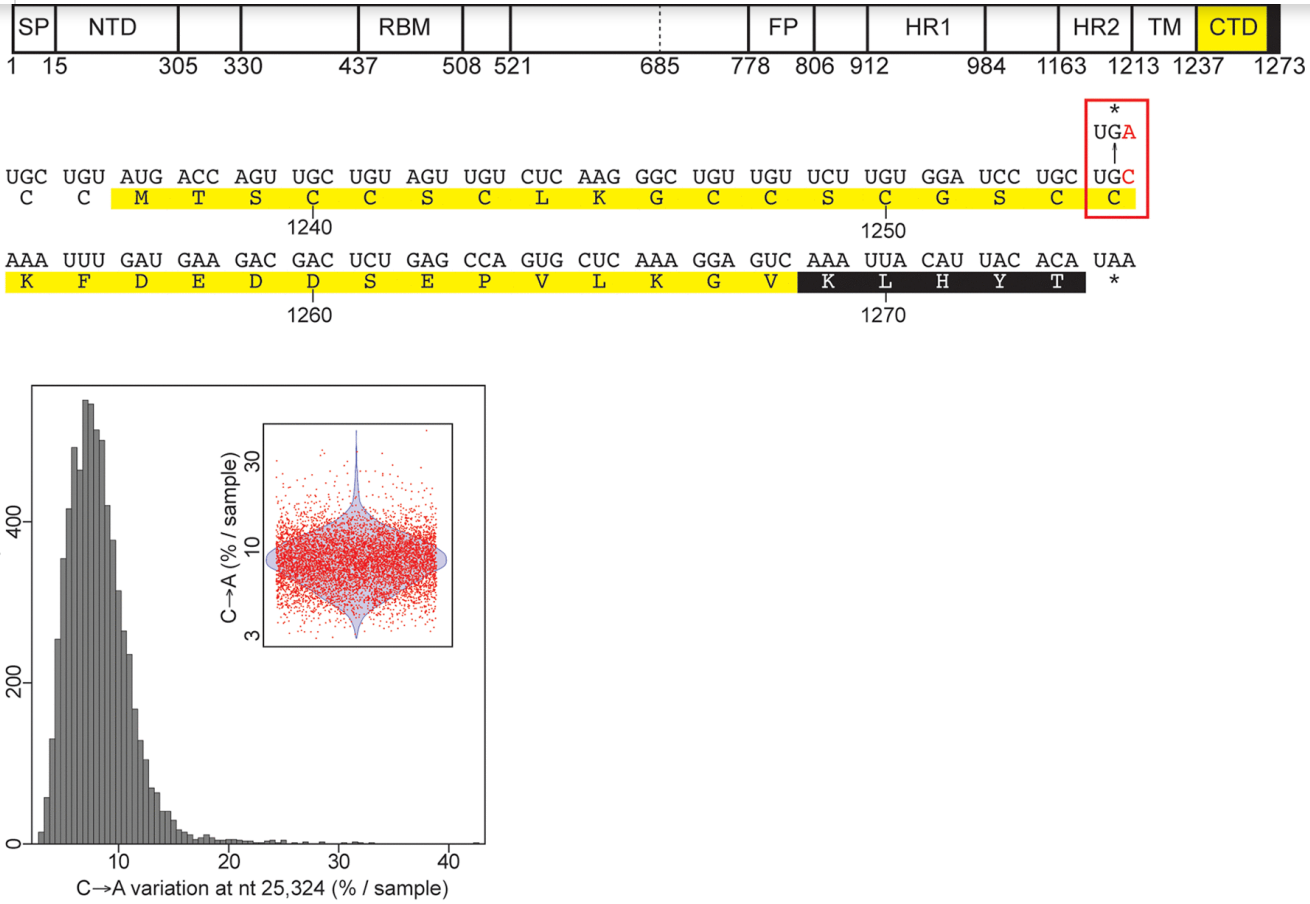


FIG 2 Localization of the C→A missense mutation on the SARS-CoV-2 S protein. (A) Schematic representation of the functional domain of the SARS-CoV-2 S protein. Below is shown the localization of the C→A variation on the carboxy-terminal domain (CTD) of the S protein. The mutation is colored and boxed in red. The carboxy-terminal domain (CTD) and the ERRS are colored in yellow and black, respectfully. (B) Distribution of the intrasample proportion of the C→A transversion at position 25,324 in the 6,668 samples containing this subspecies. The inset represents the distribution, using red dots to represent the samples having this intragenetic variation and a blue violon to show the distribution of the data.

	(%)				Consensus	Variant	Consensus	Var
29187	10.95	C→U	N	305	GCA	GUA	Ala (A)	Val
29188	10.68	A→G	N	305	GCA	GCG	Ala (A)	Ala
18591	10.21	C→G	ORF1ab	6108	GUC	GUG	Val (V)	Val
11874	10.02	U→C	ORF1ab	3870	GUA	GCA	Val (V)	Ala
15965	8.12	G→U	ORF1ab	5233	UGU	UUU	Cys (C)	Phe
29039	7.95	A→U	N	256	AAG	AUG	Lys (K)	Met (M)
6696	7.19	C→U	ORF1ab	2144	CCU	CUU	Pro (P)	Leu
28253	6.51	C→U	ORF8	120	UUC	UUU	Phe (F)	Phe
635	5.18	C→U	ORF1ab	124	CGU	UGU	Arg (R)	Cys
9502	5.17	C→U	ORF1ab	3079	GCC	GCU	Ala (A)	Ala
25323	5.14	G→C	S	1254	UGC	UCC	Cys (C)	Ser

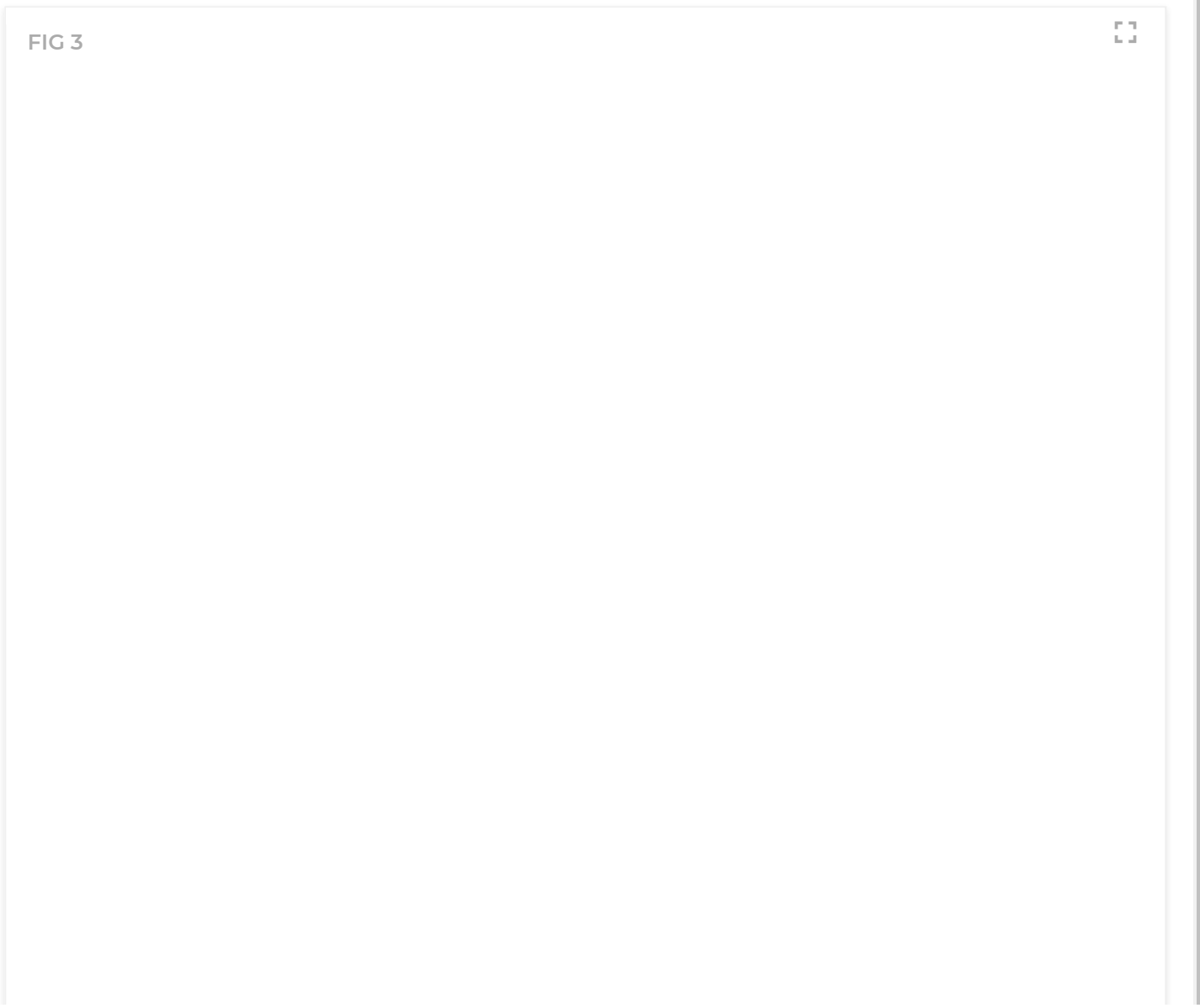
a Frequency distributions were calculated to generate statistics on the intragenetic variant populations. The variations are sorted by their recurrence, with the most shared variation at the top of the table. The mutations are indicated by underlined boldface residues.

Analysis of intragenetic variations present in SARS-CoV-2 samples from infected cells.

To further investigate variations in a more controlled system, we used 65 high-throughput sequencing data sets generated in a recent transcription profiling study of several cell lines infected with SARS-CoV-2 (29).

expected high number of A's and U's.

FIG 3



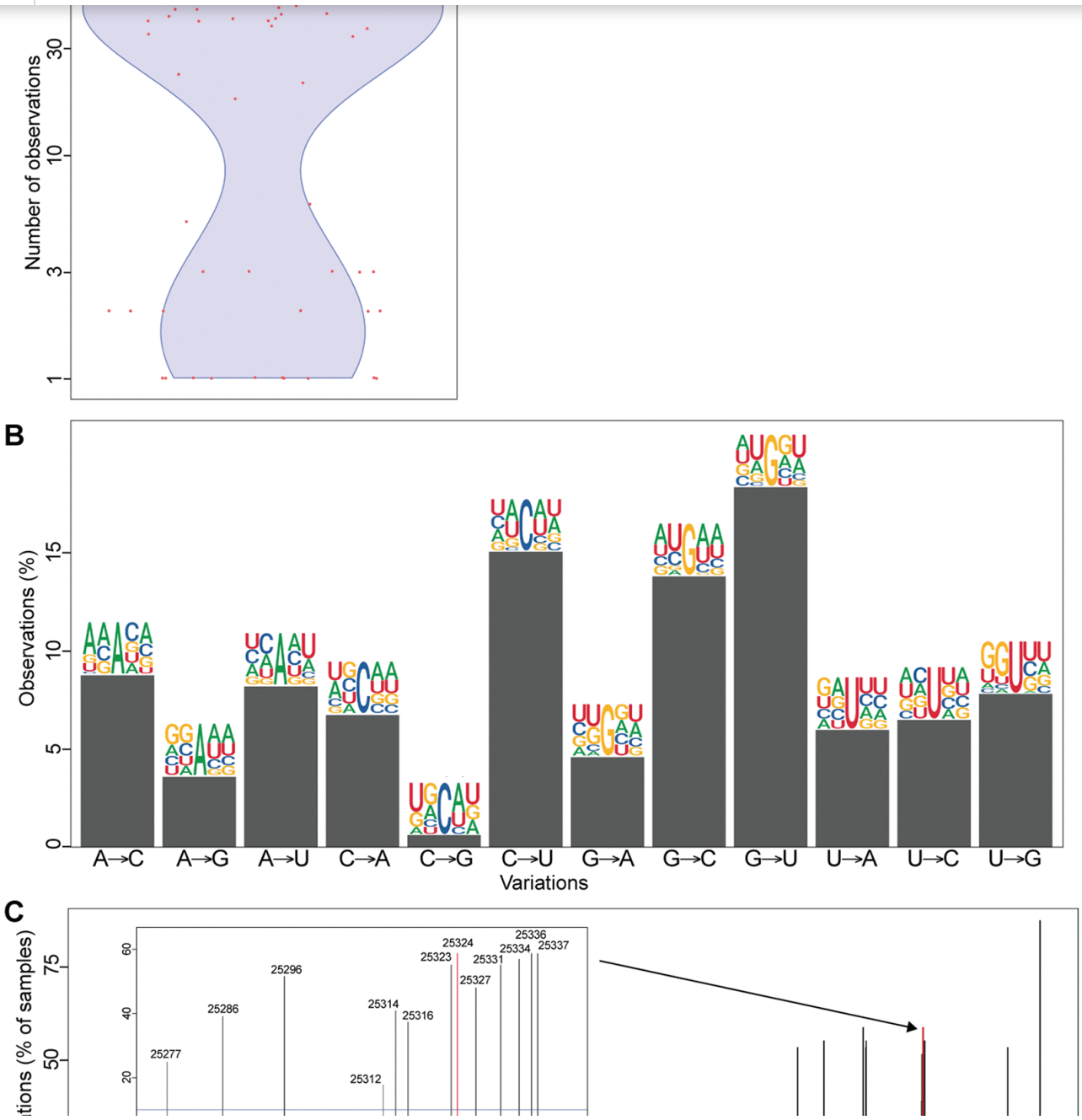


FIG 3 Intrasample variability of the SARS-CoV-2 genome in infected cells. (A) Number of intragenetic variations observed for each sample analyzed. The red dots represent the 65 samples analyzed, and the blue violon shows the distribution of the data. (B) Type of variation and sequence context for each intrasample variable position. Bars represent the percentage of each type. Sequence context is represented by logos comprised of the consensus nucleotides (center) with 2 nt upstream and 2 downstream from each intrasample variable position. (C) Recurrent intragenetic variations are represented as percentages of samples containing a variation at each position. The SARS-CoV-2 genome and its genes are represented by yellow boxes below the graph. The blue line indicates 10% shared variations and was used to extract the intrasample variations listed in Table 2. The inset represents a magnification of the cluster identified at the end of the S gene. (D) One-dimensional representation of the data shown in panel C for each type of variation individually. The location of the C→A variation at position 25,324 is indicated by a red line in panels C and D.

We then examined the intragenetic variable positions that are recurrent among the cell lines analyzed. We identified 29 positions within the viral populations showing intragenetic variation enrichment in at least 10% of the cell cultures, and most of them are located on structural genes, which are carried on the last 3′-terminal third of the viral genome (Fig. 3C and D). Similar to our observation from the samples from infected individuals, a cluster of recurrent variations is located at the 3′end of the S gene, including the C→A transversion at position 25324 shared in 58.9% of the cell lines analyzed (Fig. 3C and D, red lines; Table 2). Overall, our results indicate consistent results between intragenetic variations observed in infected cell lines and in samples from infected individuals, including the presence of the viral subspecies resulting in an S protein truncated of its last 20 amino acids (SΔ20).

	(%)				Consensus	Variant	Consensus	Var
25337	58.93	G→C	S	1259	GAC	CAC	Asp (D)	His
25334	57.14	G→C	S	1258	GAA	CAA	Glu (E)	Gln
25381	55.36	A→C	S	1273	ACA	ACC	Thr (T)	Thr
22343	55.36	G→C	S	261	GGU	CGU	Gly (G)	Arg
25323	55.36	G→C	S	1254	UGC	UCC	Cys (C)	Ser
25331	55.36	G→U	S	1257	GAU	UAU	Asp (D)	Tyr
27883	53.57	C→U	ORF7b	43	GCC	GUC	Ala (A)	Val
27882	53.57	G→C	ORF7b	43	GCC	CCC	Ala (A)	Pro
25296	51.79	A→C	S	1245	AAG	ACG	Lys (K)	Thr
23606	51.79	C→U	S	682	CGG	UGG	Arg (R)	Trp
25327	48.21	A→U	S	1255	AAA	AAU	Lys (K)	Asn
23616	48.21	G→A	S	685	CGU	CAU	Arg (R)	His
23616	44.64	G→C	S	685	CGU	CCU	Arg (R)	Pro
21550	41.07	A→C	ORF1ab	7095	AAC	CAC	Asn (N)	His
21551	41.07	A→U	ORF1ab	7095	AAC	AUC	Asn (N)	Ile
25286	39.29	A→U	S	1242	AGU	UGU	Ser (S)	Cys
25314	39.29	G→U	S	1251	GGA	GUA	Gly (G)	Val
27134	32.14	U→C	M	204	UAU	UAC	Tyr (Y)	Tyr
22206	30.36	A→G	S	215	GAU	GGU	Asp (D)	Gly
25316	30.36	U→C	S	1252	UCC	CCC	Ser (S)	Pro
26542	28.57	C→U	M	7	ACU	AUU	Thr (T)	Ile
25296	26.79	A→U	S	1245	AAG	AUG	Lys (K)	Met

	(%)				Consensus	Variant	Consensus	Var
21550	12.50	A→C	ORF1ab	7095	AAC	CAC	Asn (N)	His
21551	12.50	A→U	ORF1ab	7095	AAC	AUC	Asn (N)	Ile
25273	10.71	G→C	S	1237	AUG	AUC	Met (M)	Ile

a Frequency distributions were calculated to generate statistics on the intragenetic variant populations. The variations are sorted by their recurrence, with the most shared variation at the top of the table. The mutations are indicated by underlined boldface residues.

Increased fusogenic properties of SARS-CoV-2 SΔ20.

SARS-CoV-2 viral entry into cells is triggered by the interaction between the S glycoprotein and its cellular receptor, ACE2. While the complete mechanism of viral entry is not fully understood, it is known that S undergoes different processing steps by cellular surface and endosomal proteases. For several coronaviruses, the S protein mediates not only virion fusion but also syncytium formation (7, 12, 13). The presence of dysmorphic pneumocytes forming syncytial elements is a well-described feature of COVID-19 disease severity (30). One particularity of SARS-CoV-2 compared to SARS-CoV is the presence of an additional furin-like cleavage site at the S1/S2 interface. As a consequence, SARS-CoV-2-infected cells have a higher propensity to express activated S at the surface, which can fuse with other cells expressing the receptor ACE2 and form syncytia (30). The normal route of S trafficking involves an accumulation at the ERGIC, which is known to involve, at least in part, the interaction of the cytoplasmic portion of S with the M protein encoded by SARS-CoV-2. This interaction allows complex formation leading to virion formation at the ERGIC interface. The discovery of the SΔ20 variant missing a portion of the C terminus

fusion. Cells were lysed 24 h posttransfection, and spike processing was assessed by probing for SARS-CoV2 S1 and S2 subunits by immunoblotting. As seen in Fig. 4B and quantified in Fig. 4C, the SΔ20 protein undergoes increased processing, as observed by the presence of more S1 and S2 subunits compared to S WT (Fig. 4B, lane 2 versus lane 4). The coexpression of the M protein reduces the processing of the S WT protein while not affecting SΔ20 processing, as observed by a reduction of the S1 fragment only for the S WT (Fig. 4B, lane 3 versus lane 5). Taken together, the results shown in Fig. 4 indicate that SΔ20 displays increased processing and syncytium formation compared to the wild-type S protein and the truncation removes an important regulatory domain involving the M protein. As discussed earlier, the S protein possesses an ER retrieval signal (ERRS) at its carboxy terminus, which is required for the S protein to interact with the M protein and accumulate at the ERGIC. Deletion of this sequence in SARS-CoV was shown to reduce ERGIC accumulation within the ERGIC. We observed the same phenotype with the SARS-CoV-2 SΔ20 (Fig. 5). When M protein was coexpressed, the majority of the S WT was retained intracellularly, with little detected on the cell surface. In contrast, the majority of SΔ20 was distributed throughout the cytoplasm and on the cell surface. This result is consistent with recent observations published by Boson et al. using the SΔ19 truncation mutant (31).

FIG 4

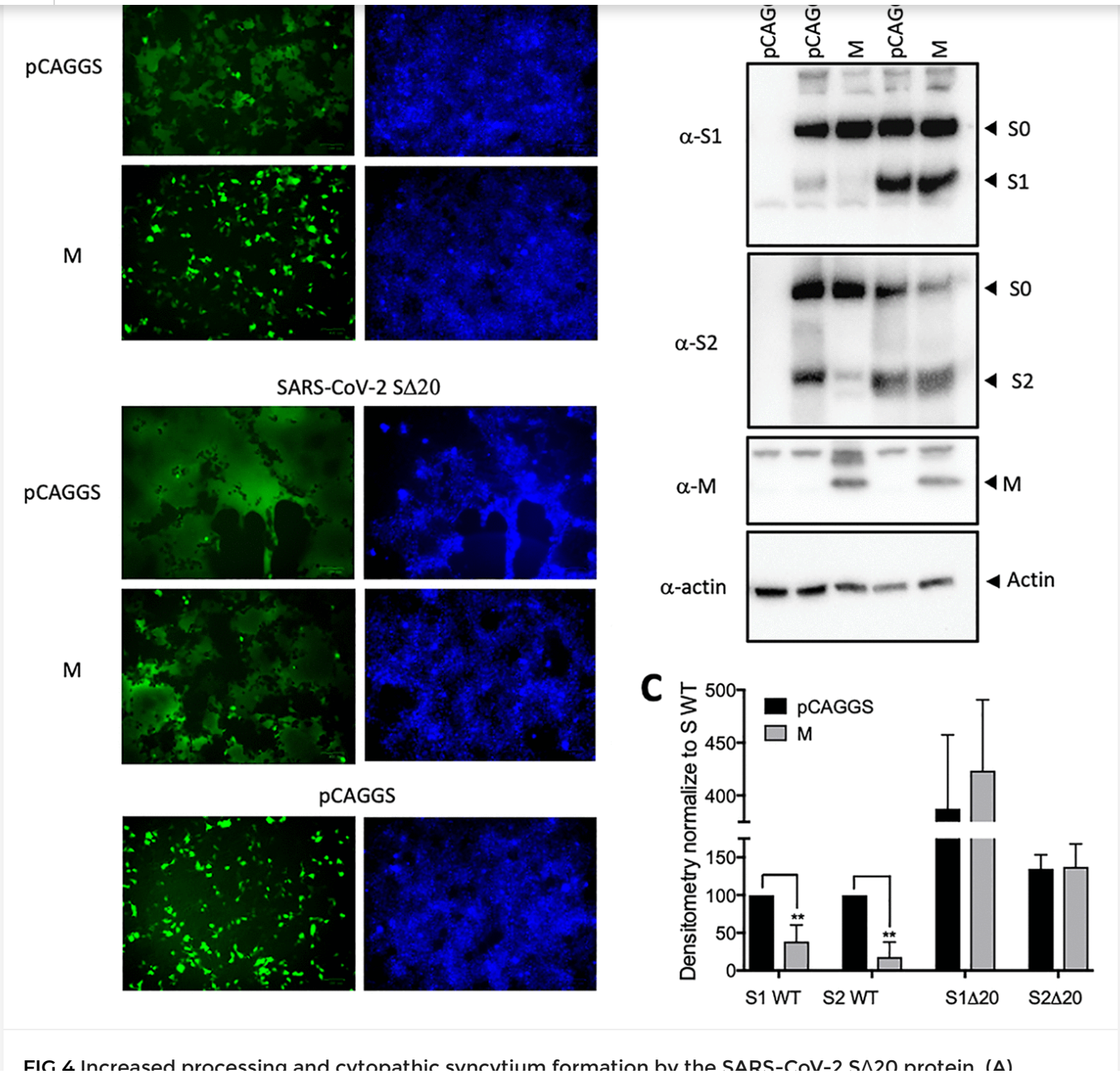


FIG 4 Increased processing and cytopathic syncytium formation by the SARS-CoV-2 SΔ20 protein. (A)

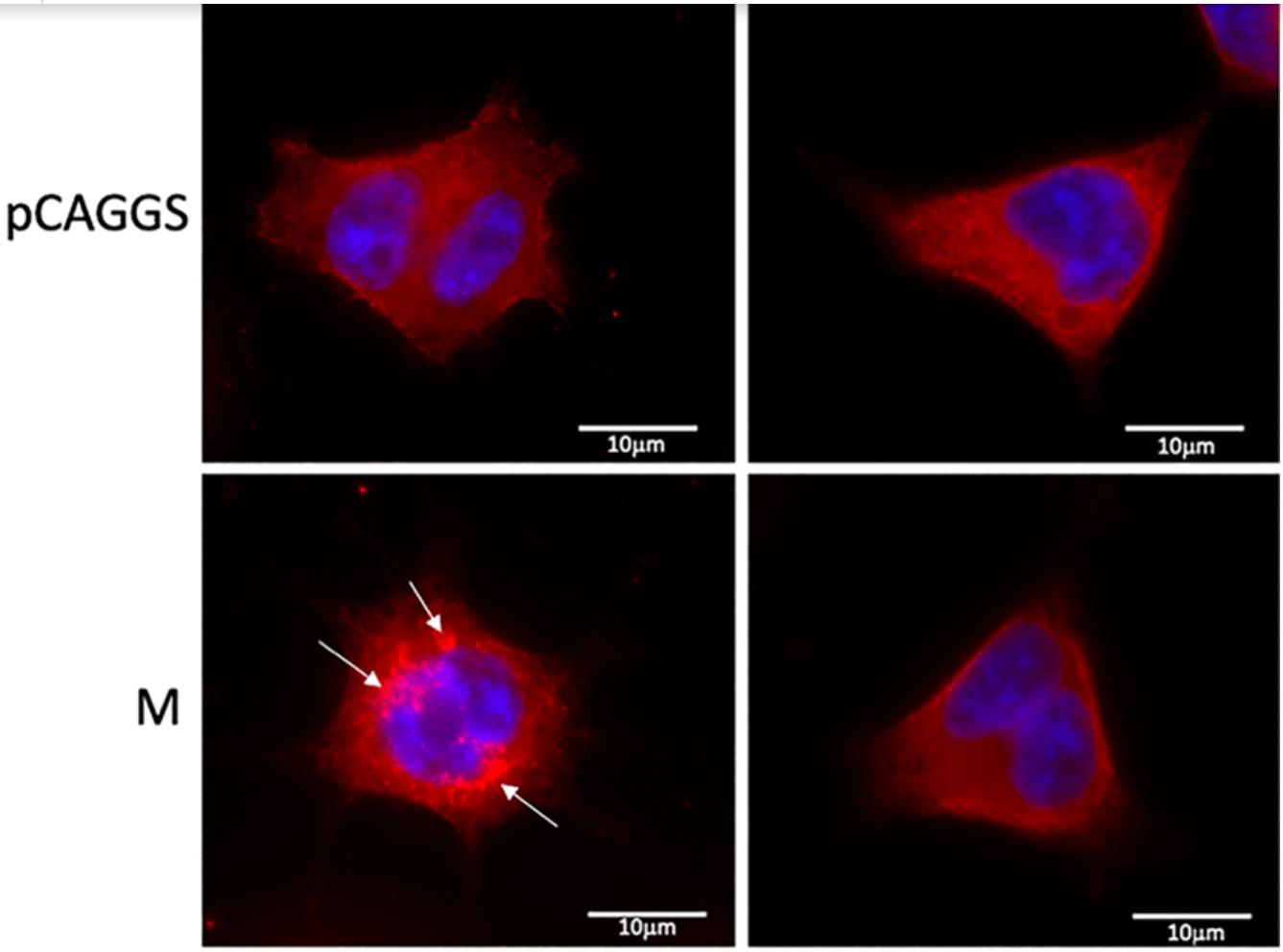


FIG 5 Subcellular localization of the S WT and SΔ20 in the presence or absence of M protein. HEK-293T cells expressing S proteins both with and without M protein were stained with anti-S protein (red), and the nucleus was stained with Hoechst 33342 dye (blue). Coexpression of M protein induced intracellular accumulation of the S WT (white arrow) but not the SΔ20 protein.

intragenetic variability in the samples analyzed. Notably, most of the samples possessed a C→A missense mutation, producing an S protein that lacks the last 20 amino acids (SΔ20) and results in increased cell-to-cell fusion and syncytium formation.

Most intrasample variations are distributed homogeneously across the viral genome and are not conserved or recurrent among samples, and a large number of them are C→U or G→U mutations. Previous analyses of SARS-CoV-2 sequence variations proposed that host editing enzymes might be involved in coronavirus transition editing, based on results showing that C→U transitions occur within a sequence context reminiscent of APOBEC1-mediated deamination (i.e., [AU]C[AU]) (21–24). Here, we investigated nucleotide compositions at each variation site and observed a high number of A's and U's around all variation types and sites. However, since the SARS-CoV-2 genome is 62% A/U-rich, and similar percentages of A's and U's were observed around all variations, we concluded that no motifs are enriched around these variations in the viral subspecies analyzed here. Consequently, our results do not allow us to conclude the frequency of intrasample genetic variations caused by host RNA editing enzymes. Previous reports used consensus sequence variation analyses to suggest the involvement of editing enzymes (21–24). If host RNA editing enzymes have a major role in coronavirus genome editing, such modified variants will likely be very abundant in the quasispecies population and thus be reflected on the consensus sequence (i.e., >50% positional frequency). In our study, the variations in each data set were compared to their respective consensus sequence. This means that if RNA editing did occur at high frequency on a defined positional hot spot, it would not have been captured by our analysis method of the quasispecies but directly reflected on the consensus sequence. We did not analyze variations in consensus sequences as this was done previously for SARS-CoV-2 (23, 24).

shearing, and/or RT-PCR amplification errors. However, we identified several positions with intrasample variability recurrent in several independent samples from both infected individuals and infected cells. They were detected at moderate to high frequencies, ranging from 2.5 to 39.3% per sample (Tables 1 and 2), and most were derived from paired-end sequencing (90.7% of the samples) in which the two strands of a DNA duplex were considered. Thus, it is likely that these variations are genuine and represent hot spots for SARS-CoV-2 genome intrasample variability.

Among the variable positions identified in infected cells, most of them are located in the last 3′-terminal third of the viral genome. These cells were infected with a large number of viruses (i.e., a high multiplicity of infection [MOI]) for 24 h (29). The presence of several variations at positions in the region coding for the main structural proteins likely reflects that this is a region with increased transcriptional activity due to the requirement of producing their encoded mRNAs from subgenomic negative-sense RNAs (8).

Interestingly, a cluster of variations located at the 3′end of the S gene was observed for the two data sets analyzed. They correspond to four transversions located at the 3′end of the S gene and are shared by a large proportion of the samples. Three of these correspond to missense mutations changing the charged side chains of two amino acids (E1258D, E1258Q, and D1259H). Notably, most of the samples possess a variability at position 25324, producing a nonsense mutation at amino acid 1254 of the S protein. The resulting protein lacks the last 20 amino acids (SΔ20) and thus does not include the ERRS motif at its carboxy terminus. For SARS-CoV-1, the ERRS domain accumulates the S protein to the ERGIC and facilitates its incorporation into virions (11). While the mechanism is not completely understood, mutation of the ERRS motif on S resulted in a failure to interact with the M protein at the ERGIC and rather resulted

Our findings indicate the presence of consistent intrasample genetic variants of SARS-CoV-2, including a recurrent subpopulation of SΔ20 variants with elevated fusogenic properties. It is tempting to suggest a link between SARS-CoV-2 pathogenesis and the presence of SΔ20, since severe cases of the disease were recently linked to considerable lung damage and the occurrence of syncytia (30, 49). Also, as observed for several enveloped viruses, syncytium formation could allow cell-to-cell spreading without virion production, which could facilitate not only viral dissemination but also immune evasion (50). Clearly, more investigation is required to better define the extent of SARS-CoV-2 variability in infected hosts and to assess the role of these subspecies in the life cycle of this virus. More importantly, further studies on the presence of SΔ20 and its link with viral pathogenicity could lead to better diagnostic strategies and design treatments for COVID-19.

MATERIALS AND METHODS

Analysis of intragenetic variability within SARS-CoV-2 samples.

A total of 15,289 publicly available high-throughput sequencing data sets were downloaded from the NCBI Sequence Read Archive (up to 10 July 2020). They comprise 15,224 data sets from infected individuals and 65 data sets from infected cell lines. Table S1 in the supplemental material includes all of the accession numbers. All data sets were derived from Illumina sequencing technology. The data sets from infected cells were generated by Blanco-Melo et al. (29). Duplicated reads were combined to reduce amplification bias and mapped to the SARS-CoV-2 isolate Wuhan-Hu-1 reference genome (NC_045512v2) using hisat2

the NCBI Sequence Read Archive (SRA) and used in this study. Only SRA numbers are listed. The 65 data sets from infected cells also include "(cells)" as a label and are listed at the end.
Download Table S1, TXT file, 0.2 MB.

Cell culture and plasmids.

Human embryonic kidney 293T (HEK-293T) cells were obtained from the American Type Culture Collection (ATCC CRL-11268) and maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal bovine serum (Fisher Scientific), 5% bovine calf serum (Fisher Scientific), 100 U/ml penicillin, and 100 μg/ml streptomycin (Fisher Scientific). HEK-293T cells stably expressing human ACE2 (HEK-293T-hACE2 cell line; BEI Resources) were cultured and maintained in DMEM (Corning) supplemented with 10% fetal bovine serum (Sigma), 100 U/ml penicillin, and 100 μg/ml streptomycin. All cells were cultured at 37°C in a humidified atmosphere containing 5% CO_2. pCAGGS expressing the SARS-CoV-2 S protein (Wuhan-Hu-1; WT) was provided by Florian Krammer (Mount Sinai). SARS-CoV-2 SΔ20 was generated using overlapping PCR to introduce a termination codon at residue 1254. The expression construct encoding SARS-CoV-2 M was generated by PCR amplification of the M gene from pLVX-EF1alpha-SARS-CoV-2-M-2×Strep-IRES-Puro (a kind gift of Nevan Krogan, UCSF) and addition of a stop codon to remove the Strep (streptavidin) tag prior to cloning into pCAGGS.

Syncytium formation assay.

Twenty-four-well plates were seeded with HEK-293T-hACE2 cells in complete medium to obtain 90%

with cold phosphate-buffered saline (PBS) and lysed in cold lysis buffer (1% Triton X-100, 0.1% IGEPAL CA-630, 150 mM NaCl, 50 mM Tris-HCl, pH 7.5) containing protease and phosphatase inhibitors (Cell Signaling). Proteins in cell lysates were resolved on 4 to 12% gradient SDS-polyacrylamide gels (NuPage; Invitrogen) and transferred to polyvinylidene difluoride (PVDF) membranes. Membranes were blocked for 1 h at room temperature with blocking buffer (5% skim milk powder dissolved in 25 mM Tris, pH 7.5, 150 mM NaCl, and 0.1% Tween 20 [TBST]). Processing of spike protein was detected by immunoblotting using an anti-S1 antibody (SARS-CoV/SARS-CoV-2 spike protein S1 polyclonal; Invitrogen) and anti-S2 antibody (SARS-CoV/SARS-CoV-2 spike protein S2 monoclonal; Invitrogen). Overexpression of M was also detected by immunoblotting and using an anti-M antibody (rabbit anti-SARS membrane protein; Novus Biologicals). Membranes were incubated overnight at 4°C with the appropriate primary antibody in the blocking buffer. Blots were then washed in TBST and incubated with horseradish peroxidase (HRP)-conjugated secondary antibody for 1 h at room temperature (anti-mouse HRP and anti-rabbit HRP; both from Cell Signaling). Membranes were washed, incubated in chemiluminescence substrate (SuperSignal West Femto Maximum Sensitivity substrate; Thermo Fisher Scientific), and imaged using the ChemiDoc XRS+ imaging system (Bio-Rad). In some instances, the same membrane was stripped and reprobed for actin (monoclonal anti-β-actin; Millipore Sigma). Densitometry was performed using ImageJ software (56) and data analysis with Prism 8 (GraphPad).

Immunofluorescence.

HEK-293T cells were transiently cotransfected using JetPRIME (Polyplus Transfection, France) with plasmids encoding SARS-CoV-2 S or SARS-CoV-2 SΔ20 and M proteins. Twenty-four hours

Chair in Molecular Virology and Intrinsic Immunity. M.C. is a Canada Research Chair in Molecular Virology and Antiviral Therapeutics. This work was supported by a COVID-19 Rapid Research grant from the Canadian Institutes for Health Research (CIHR; OV1 170355) to M.-A.L. and M.P. and a COVID-19 Rapid Research Grant (OV3 170632) to M.C. and P.M.G.

REFERENCES

1. Hu B, Guo H, Zhou P, Shi Z-L. 2020. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19:141–154.
Go to Citation | Crossref | PubMed | Google Scholar

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733.
Go to Citation | Crossref | PubMed | Google Scholar

3. Fehr AR, Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* 1282:1–23.
Go to Citation | Crossref | PubMed | Google Scholar

4. Hartenian E, Nandakumar D, Lari A, Ly M, Tucker JM, Glaunsinger BA. 2020. The molecular virology of coronaviruses. *J Biol Chem* 295:12910–12934.
Go to Citation | Crossref | PubMed | Google Scholar

5. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. 2020. A structural view of SARS-CoV-2

8. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. 2020. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 19:155–170.
Go to Citation | Crossref | PubMed | Google Scholar

9. McBride CE, Li J, Machamer CE. 2007. The cytoplasmic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. *J Virol* 81:2418–2428.
Go to Citation | Crossref | PubMed | Google Scholar

10. Westerbeck JW, Machamer CE. 2019. The infectious bronchitis coronavirus envelope protein alters Golgi pH to protect the spike protein and promote the release of infectious virus. *J Virol* 93:e00015-19.
Go to Citation | Crossref | PubMed | Google Scholar

11. Lontok E, Corse E, Machamer CE. 2004. Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. *J Virol* 78:5913–5922.
Go to Citation | Crossref | PubMed | Google Scholar

12. Qian Z, Dominguez SR, Holmes KV. 2013. Role of the spike glycoprotein of human Middle East respiratory syndrome coronavirus (MERS-CoV) in virus entry and syncytia formation. *PLoS One* 8:e76469.
Go to Citation | Crossref | PubMed | Google Scholar

13. Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. 2010. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J Virol* 84:12658–12664.
Go to Citation | Crossref | PubMed | Google Scholar

16. Phan T. 2020. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol* 81:104260.
Go to Citation | Crossref | PubMed | Google Scholar

17. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT, Balloux F. 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* 83:104351.
Go to Citation | Crossref | PubMed | Google Scholar

18. Vankadari N. 2020. Overwhelming mutations or SNPs of SARS-CoV-2: a point of caution. *Gene* 752:144792.
Go to Citation | Crossref | PubMed | Google Scholar

19. Mavian C, Marini S, Prosperi M, Salemi M. 2020. A snapshot of SARS-CoV-2 genome availability up to April 2020 and its implications: data analysis. *JMIR Public Health Surveill* 6:e19170.
Go to Citation | Crossref | PubMed | Google Scholar

20. Farkas C, Fuentes-Villalobos F, Garrido JL, Haigh J, Barría MI. 2020. Insights on early mutational events in SARS-CoV-2 virus reveal founder effects across geographical regions. *PeerJ* 8:e9255.
Go to Citation | Crossref | PubMed | Google Scholar

21. Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat Struct Mol Biol* 18:230–236.
Go to Citation | Crossref | PubMed | Google Scholar

25. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol* 84:9733–9748.
Go to Citation | Crossref | PubMed | Google Scholar

26. Drake JW, Holland JJ. 1999. Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* 96:13910–13913.
Go to Citation | Crossref | PubMed | Google Scholar

27. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. 2011. Coronaviruses. *RNA Biol* 8:270–279.
Go to Citation | Crossref | PubMed | Google Scholar

28. Domingo E, Perales C. 2019. Viral quasispecies. *PLoS Genet* 15:e1008271.
Go to Citation | Crossref | PubMed | Google Scholar

29. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA, tenOever BR. 2020. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 181:1036–1045.e9.
Go to Citation | Crossref | PubMed | Google Scholar

30. Bussani R, Schneider E, Zentilin L, Collesi C, Ali H, Braga L, Volpe MC, Colliva A, Zanconati F, Berlot G, Silvestri F, Zacchigna S, Giacca M. 2020. Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology. *EBioMedicine* 61:103104.
Go to Citation | Crossref | PubMed | Google Scholar

Go to Citation | Crossref | PubMed | Google Scholar

34. Chen L, Liu P, Evans TC, Ettwiller LM. 2017. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* 355:752–756.
Go to Citation | Crossref | PubMed | Google Scholar

35. Belhadj Slimen I, Najar T, Ghram A, Dabbebi H, ben Mrad M, Abdrabbah M. 2014. Reactive oxygen species, heat stress and oxidative-induced mitochondrial damage. A review. *Int J Hyperthermia* 30:513–523.
Go to Citation | Crossref | PubMed | Google Scholar

36. Ahn EH, Lee SH. 2019. Detection of low-frequency mutations and identification of heat-induced artifactual mutations using duplex sequencing. *Int J Mol Sci* 20:199.
Go to Citation | Crossref | Google Scholar

37. Arbeithuber B, Makova KD, Tiemann-Boege I. 2016. Artifactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications. *DNA Res* 23:547–559.
Go to Citation | Crossref | PubMed | Google Scholar

38. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* 108:9530–9535.
Go to Citation | Crossref | PubMed | Google Scholar

39. Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA. 1992. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. *J Biol Chem* 267:166–172.
Go to Citation | Crossref | PubMed | Google Scholar

Go to Citation | Crossref | PubMed | Google Scholar

43. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 11:1620.
Go to Citation | Crossref | PubMed | Google Scholar

44. Dieterle ME, Haslwanter D, Bortz RH, Wirchnianski AS, Lasso G, Vergnolle O, Abbasi SA, Fels JM, Laudermilch E, Florez C, Mengotto A, Kimmel D, Malonis RJ, Georgiev G, Quiroz J, Barnhill J, Pirofski L, Daily JP, Dye JM, Lai JR, Herbert AS, Chandran K, Jangra RK. 2020. A replication-competent vesicular stomatitis virus for studies of SARS-CoV-2 spike-mediated cell entry and its inhibition. *Cell Host Microbe* 28:486–496.e6.
Go to Citation | Crossref | PubMed | Google Scholar

45. Schmidt F, Weisblum Y, Muecksch F, Hoffmann H-H, Michailidis E, Lorenzi JCC, Mendoza P, Rutkowska M, Bednarski E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Caskey M, Robbiani DF, Nussenzweig MC, Rice CM, Hatziioannou T, Bieniasz PD. 2020. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J Exp Med* 217:e20201181.
Go to Citation | Crossref | PubMed | Google Scholar

46. Case JB, Rothlauf PW, Chen RE, Liu Z, Zhao H, Kim AS, Bloyet L-M, Zeng Q, Tahan S, Droit L, Ilagan MXG, Tartell MA, Amarasinghe G, Henderson JP, Miersch S, Ustav M, Sidhu S, Virgin HW, Wang D, Ding S, Corti D, Theel ES, Fremont DH, Diamond MS, Whelan SPJ. 2020. Neutralizing antibody and soluble ACE2 inhibition of a replication-competent VSV-SARS-CoV-2 and a clinical isolate of SARS-CoV-2. *Cell Host Microbe* 28:475–485.e5.
Go to Citation | Crossref | PubMed | Google Scholar

47. Stephens EB, Compans RW. 1988. Assembly of animal viruses at cellular membranes. *Annu*

Go to Citation | Crossref | Google Scholar

51. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907–915.
Go to Citation | Crossref | PubMed | Google Scholar

52. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
Go to Citation | Crossref | PubMed | Google Scholar

53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25:2078–2079.
Go to Citation | Crossref | PubMed | Google Scholar

54. Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics (Oxford, England)* 33:3645–3647.
Go to Citation | Crossref | PubMed | Google Scholar

55. Côté M, Zheng Y-M, Liu S-L. 2009. Receptor binding and low pH coactivate oncogenic retrovirus envelope-mediated fusion. *J Virol* 83:11447–11455.
Go to Citation | Crossref | PubMed | Google Scholar

56. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671–675.
Go to Citation | Crossref | PubMed | Google Scholar

Related ArticlesRelated Articles From ASM